DOCUMENT RESUME

ED 117 152		TM 005 006

AUTHOR Huberty, Carl J.; Curry, Allen R.
TITLE Linear Versus Quadratic Multivariate Classification.
PUB DATE [Apr 75]
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS *Analysis of Covariance; Bayesian Statistics; *Classification; *Comparative Analysis; *Discriminant Analysis; Matrices; Predictor Variables; Probability; *Sampling; Statistical Analysis; Tests of Significance
IDENTIFIERS Linear Classification; Quadratic Classification

ABSTRACT

A linear classification rule (used with equal covariance matrices) was contrasted with a quadratic rule (used with unequal covariance matrices) for accuracy of internal and external classification. The comparisons were made for seven situations which resulted from combining three data conditions (equal and unequal covariance matrices, minimal and nonminimal group centroid separation, and two and three criterion groups) for different sets of data. For the internal analysis the quadratic rule was superior in all seven situations. For the external analysis the linear rule was superior in nearly all of the situations. (Author)

ED117152

Linear Versus Quadratic

Multivariate Classification

Carl J Huberty and Allen R. Curry

University of Georgia

TM005 006

Linear Versus Quadratic Multivariate Classification

Introduction

Multivariate classification may be considered as one aspect of discriminant analysis -- other aspects being separation, discrimination, and estimation. A classification analysis is primarily applicable to the following problem: Given p measures associated with an individual (or object), can we predict the one of K well-defined and exhaustive populations to which this individual most likely belongs? Classification serves other potentially useful purposes as well. For example, the proportion of correct classifications (or assignments) may be used as in index of discriminatory power of a set of predictors. Results of a classification analysis may also be used for assessing the relative contribution of the predictors to criterion population separation.

Various multivariate classification rules have been proposed. Although some nonparametric rules have been advanced, most research dealing with the study and application of rules has involved those rules that are parametric in nature. In particular, rules based on multivariate normal distribution theory have been the most popular. One criterion for selecting a class of appropriate rules from those available is the similarity of covariance structure of the predictors across the K criterion populations. If it can be assumed, or if the sample data suggest, that the covariance structure is the same, a "linear" rule is selected; if not, a nonlinear rule would be the choice. If it is decided that a nonlinear rule would be appropriate, the choice has typically been a "quadratic" rule. (See Huberty, in press, for elaboration.)

The equal covariance structure condition has typically been ignored in applications of multivariate classification, with one or another linear rule being employed. The question then arises as to whether or not some predictive accuracy has been lost when a linear (or quadratic) rule rather than a quadratic (or linear) rule has been used. The purpose of the present investigation was to compare the accuracy of a linear classification rule with that of a quadratic rule. The comparison was made under the conditions considered appropriate for the use of each type of rule.

Data

Three data conditions were considered in combination to yield eight "situations." The first condition deals with the equality or inequality of the predictor variable population covariance matrices; this condition was assessed via a test proposed by G.E.P. Box, which is a generalization of the Bartlett test for the homogeneity of K univariate variances (see Cooley and Lohnes, 1971, p. 229). The second condition is that of the degree of seperation of the K population centroids (or mean vectors), as assessed by Wilks's lambda criterion (see Cooley and Lohnes, 1971, p. 226). (The Wilks's criterion was employed recognizing that its appropriateness depends, strictly, upon the condition of equal covariance matrices). The third condition is the number of criterion groups studied.

As mentioned above, when there is insufficient evidence to conclude that the covariance matrices are unequal, a linear rule is generally considered appropriate. Under this condition, a linear rule was contrasted with a quadratic rule for minimal and nonminimal centroid separation for two and three criterion groups -- four situations resulting. When the data suggested that the covariance matrices were unequal, the two rules were again contrasted for the four situations.

Three data sets were employed for the comparisons. Within Set A the subjects are public school reading teachers: the 10 predictor measures used are measures of knowledge of reading and of teacher background; the criterion groups are defined by method of reading instruction employed. Data Set B is based on college freshmen: measures on high school academic performance, standardized tests of French achievement, and nationally normed tests for college bound students provide scores on 13 predictors; the criterion groups are defined by instructor judgment of student placement in college French classes. The subjects of data Set C are high school students: the 17 predictor measures are cognitive, interest, personality, and socioeconomic status measures; criterion groups are based upon post-secondary educational placement. To provide data that indicated equal covariance structures, complete groups were deleted from each data set, retaining unequal group sizes. A situation with three criterion groups that are minimally separated, for which a linear classification rule would be appropriate, was not investigated, since data for such a situation were unavailable. Thus, seven of eight possible data situations were considered.

Data Analysis

The linear classification rule used in this study is based on a Bayesian conditional-probability model assuming multivariate normality within each criterion population, and constant covariance structure across the criterion populations. The classification statistic is a function of sample mean vectors and the within-groups covariance matrix. Defining

$$D_{ik}^2 = (\underline{X}_i - \bar{\underline{X}}_k)' S^{-1} (\underline{X}_i - \bar{\underline{X}}_k)$$

to be square of the distance from the point in p-space representing individual $i(\bar{\underline{X}}_i)$ to the point representing the means of the p measures in group k $(\bar{\underline{X}}_k)$, where S is the pooled sample (pxp) covariance matrix, the following classification statistic was used:

$$P_{ik} = \frac{p_k \exp(-\frac{1}{2}D_{ik}^2)}{\sum_{k'=1}^{K} p_{k'} \exp(-\frac{1}{2}D_{ik'}^2)},$$

where p_k is the prior probability of membership in population k. This latter expression represents the (posterior) probability of individual i belonging to population k. An individual is classified into that population from which the sample yields the largest value of P_{ik}. The value of p_k used in this study is N_k/N, where N_k is the size of the sample selected from population k, and $N = \sum_k N_k$.

The quadratic classification rule used is similar to the linear rule except that the sample covariance matrix for each group (S_k) is used in place of S, with the determinants of the S_k matrices incorporated (see Cooley & Lohnes, 1971, p. 268).

In comparing the accuracy of prediction of the linear rule to that of the quadratic rule, both "internal" and "external" classification results were considered. Results of an internal classification analysis are those obtained when measures for the individuals on whom the statistics ($\bar{\underline{X}}_k$ and S or S_k) were based are resubstituted to obtain the P_{ik} values. In an external classification analysis statistics based on one set of individuals are used in classifying "new" individuals. The external classification method used in this study is an extension of that suggested by Lachenbruch

(1967). The procedure for the Lachenbruch method is as follows: Compute the statistics for each of the possible total samples of size $\sum_k N_k - 1$ obtained by omitting one individual's vector from the original total sample, and record for each computation whether the omitted individual is misclassified. In calculating the P_{ik} values for both the linear and quadratic rules, matrix inversions are requried, but the labor can be reduced to merely adjusting the inverses based on all $\sum_k N_k$ individuals. Expressions[1] for the adjustments of S^{-1}, S_k^{-1}, and the mean vectors are given by Eisenbeis and Avery (1972, p. 100).

Separate group as well as total group proportions of correct classifications were compared for the linear and quadratic rules; McNemar's chi-square statistic was used in the statistical comparisons of the total sample proportions. Measures of distances (Mahalanobis D^2 with modifications for unequal covariance matrices) in multivariate spaces between pairs of group mean vectors were examined to determine group proximity. An "arrant misclassification" is defined as one that occurs when if an individual is misclassified, he is classified into a population other than one "closest" to his actual population. The two rules were compared in terms of the number of arrant misclassifications for both the internal and external analyses.

Results

Means, standard deviations, univariate ANOVA mean-square ratios, and within-groups intercorrelations of the predictors were determined for each

situation. Tables of such values are available upon request.

The seven data situations investigated were characterized by (a) number of criterion groups, (b) group separation, and (c) the appropriate, in terms of covariance structure, classification rule (see Table 1). A situation with minimal separation was arbitrarily defined to be one for which $\Lambda > .80$; for nonminimal separation, $\Lambda \leq .80$. Thus, in situations I, IV, and V the groups are minimally separated. If the F statistic used to test the equality of the population covariance matrices yielded significance ($p < .05$) a quadratic rule was judged appropriate, otherwise a linear rule was considered appropriate.

Insert Table 1 about here

The results of the internal and external classification analysis for the linear and quadratic rules are reported in Tables 2 through 8. The expected proportions given in the tables are based on the marginal sums for each classification matrix. The groups are listed in the "order" determined by the multivariate distance measures. Various results are clear from the tables. First, consider a comparison of the linear and the quadratic rules for the internal analysis. The proportion of correct classifications across all criterion groups is significantly higher for the quadratic rule than for the linear rule in all situations -- the smallest value of McNemar's chi-square statistic was 5.76 with $p<.025$. And with two exceptions the quadratic rule outperforms the linear rule in terms of proportions of correct classifications for separate groups. One exception is for situation V (see Table 6) where the proportion with the linear rule for group 1 ($53/65 = 0.82$) is slightly higher than that with the quadratic rule for group 1 ($51/65 = 0.78$) -- note that group 1 is the largest group. The other

exception is for situation VII (see Table 8) where the group 3 proportion with the linear rule (161/200 = 0.805) is about the same as that with the quadratic rule (159/200 = 0.795); identical proportions resulted for group 1 -- note again that groups 3 and 1 are the largest groups. The number of arrant misclassifications (appropriately considered only in situations in which three groups were involved) was less with the quadratic rule in situation V (see Table 6, 21 versus 31) and in situation VII (see Table 8, 58 versus 67).

Insert Tables 2-8 about here

Second, consider a comparison of the two rules for the external analysis. For situations I, IV, and V (see Tables 2, 5, and 6) the across-group proportions were about what would be expected by chance classification, for the given marginal sums; note that for all three of these situations the group separation was minimal. For situations II and III (see Tables 3 and 4) in which the linear rule was judged appropriate and separation was nonminimal, the linear rule did better with the difference being statistically significant ($p<.05$) for situation III. The linear rule also gave better results (.648 versus .604, $p<.05$) for situation VII (see Table 8), where the quadratic rule was appropriate and separation was nonminimal. For situation VI (see Table 7) where the quadratic rule was appropriate and separation was nonminimal, the quadratic rule was clearly better ($p<.001$). For all situations but one, the proportion of correct classifications for the largest group was highest with the linear rule; the exception was situation VII where the quadratic rule yielded 87.6% correct classifications while the linear rule yielded 84.5%. The linear rule also yielded fewer arrant misclassifications for situation V (see Table 6, 21 versus 37), while the numbers were identical for situation III (see Table 4), and nearly the same for situation VII (see Table 8).

Third, consider a comparison of the internal analysis and the external

analysis. As to be expected, the internal analysis yielded higher proportions of correct classifications than the external analysis in all situations save one for both rules. The lone exception was for situation VI (see Table 7) where the proportions were identical with the quadratic rule.

Discussion

If, in a study calling for a multivariate classification analysis, interest is primarily on obtaining a high proportion of correct classifications in an "internal" sense, then a quadratic rule should always be used in preference to a linear rule. With this concern the quadratic rule would be used regardless of the covariance structure of the data. However, if the concern is for high classification accuracy for a new data set (i.e., "external" classification), then, based on the results of the current investigation, a quadratic rule should not always be used. It was found that the linear rule yielded a higher across-group proportion of correct classifications for an external analysis for two situations involving three criterion groups that have nonminimal separation. That a linear rule did better than a quadratic rule in an external sense is presumably due to the fact that fewer parameters need be estimated with the linear rule. It is conjectured that the results of an external analysis would be improved if only the "better" predictors were used in the analysis. (This conjecture was supported by the results of an external analysis of the data of situation VII with only nine of the predictor measures used. The results are given in Table 9.) With regard to separate group classification accuracy,

---

Insert Table 9 about here

---

based on the results of this study, it might be recommended that a linear rule be used when interest is mainly on getting high accuracy for the largest criterion group.

Whereas proportions of correct classifications obtained from an internal classification analysis are known to constantly overestimate the true proportions (i.e., probabilities), external classification gives an underestimation. The difference between proportions yielded by the two analyses indicates the interval in which the "optimal probability" can be expected to lie. If there is a great difference between the two proportions, one can expect to achieve better classification of new samples by increasing sample sizes (Michaelis, 1973, p. 233).

The present investigation represents only a beginning. More empirical investigations are needed in the study of linear versus quadratic classification, using both internal and external analyses. Perhaps some Monte Carlo studies are called for, taking into consideration such factors as covariance structure, number of predictors, sample sizes, group separation, and predictor intercorrelations, to list a few.

References

Cooley, W. W. & Lohnes, P. R. Multivariate data analysis. New York: Wiley, 1971.

Eisenbeis, R. A., & Avery, R. B. Discriminant analysis and classification procedures. Lexington, Mass.: Heath, 1972.

Huberty, C. J. Discriminant analysis. Review of Educational Research, in press.

Lachenbruch, P. A. An almost unbieased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Biometrics, 1967, 23, 657-662.

Michaelis, J. Simulation experiments with multiple group linear and quadratic discriminant analysis. In T. Cacoullos (Ed.), Discriminant analysis and applications. New York: Academic Press, 1973.

Footnote

[1]The Eisenbeis and Avery expressions for the adjustments of S^{-1} and of S_k^{-1} are in error. In each, the first sign within the brackets should be plus rather than minus.

Table 1

Description of the Seven

Data Situations Investigated

Situation	Number of Groups	Number of Variables	Sample Sizes	Wilks's Lambda	F- and df-values for Equality of Covariance Matrices	Appropriate Rule
I	2	10	65,47	.9464	1.096; df:55, ∞[a]	Linear
II	2	13	35,81	.3583	1.043; df:91, ∞	Linear
III	3	13	35,81,37	.2313	1.152; df:182,∞	Linear
IV	2	10	65,40	.8923	1.589; df:55, ∞	Quadratic
V	3	10	65,47,40	.9119	1.278; df:110,∞	Quadratic
VI	2	17	26,200	.7672	1.431; df:153,∞	Quadratic
VII	3	17	177,75,200	.5509	1.650; df:306,∞	Quadratic

[a]df value greater than 10,000.

Table 2

Frequencies of Classifications

for Situation I

(Equal Covariances, Minimal Separation, Two Groups)

Internal Classification

Linear

	Classified Group 1	2	Total
Actual 1	56	9	65
Group 2	31	16	47

$P_o = .643$
$P_e = .544$

Quadratic

	Classified Group 1	2	Total
Actual 1	57	8	65
Group 2	19	28	47

$P_o = .759$
$P_e = .529$

External Classification

Linear

	Classified Group 1	2	Total
Actual 1	47	18	65
Group 2	39	8	47

$P_o = .491$
$P_e = .543$

Quadratic

	Classified Group 1	2	Total
Actual 1	43	22	65
Group 2	34	13	47

$P_o = .500$
$P_e = .531$

P_o = observed proportion of correct classifications across all groups.

P_e = expected proportion of correct classifications across all groups.

Table 3

Frequencies of Classifications

for Situation II

(Equal Covariances, Nonminimal Separation, Two Groups)

Internal Classification

Linear

		Classified Group 1	2	Total
Actual	1	30	5	35
Group	2	5	76	81

$P_o = .914$
$P_e = .579$

Quadratic

		Classified Group 1	2	Total
Actual	1	33	2	35
Group	2	1	80	81

$P_o = .974$
$P_e = .582$

External Classification

Linear

		Classified Group 1	2	Total
Actual	1	29	6	35
Group	2	7	74	81

$P_o = .888$
$P_e = .575$

Quadratic

		Classified Group 1	2	Total
Actual	1	23	12	35
Group	2	7	74	81

$P_o = .836$
$P_e = .596$

Table 4

Frequencies of Classifications

for Situation III

(Equal Covariances, Nonminimal Separation, Three Groups)

Internal Classification

Linear

		Classified Group			
		1	2	3	Total
Actual	1	30	5	0	35
Group	2	7	71	3	81
	3	0	5	32	37

$P_o = .869$
$P_e = .391$

Quadratic

		Classified Group			
		1	2	3	Total
Actual	1	33	2	0	35
Group	2	1	77	3	81
	3	0	3	34	37

$P_o = .941$
$P_e = .393$

External Classification

Linear

		Classified Group			
		1	2	3	Total
Actual	1	29	6	0	35
Group	2	7	70	4	81
	3	0	8	29	37

$P_o = .837$
$P_e = .397$

Quadratic

		Classified Group			
		1	2	3	Total
Actual	1	23	12	0	35
Group	2	7	68	6	81
	3	0	12	25	37

$P_o = .758$
$P_e = .411$

Table 5

Frequencies of Classifications

for Situation IV

(Unequal Covariances, Minimal Separation, Two Groups)

Internal Classification

Linear

	Classified Group 1	2	Total
Actual 1	55	10	65
Group 2	24	16	40

P_o = .676
P_e = .560

Quadratic

	Classified Group 1	2	Total
Actual 1	56	9	65
Group 2	14	26	40

P_o = .781
P_e = .540

External Classification

Linear

	Classified Group 1	2	Total
Actual 1	52	13	65
Group 2	27	13	40

P_o = .619
P_e = .560

Quadratic

	Classified Group 1	2	Total
Actual 1	44	21	65
Group 2	26	14	40

P_o = .552
P_e = .540

Table 6

Frequencies of Classifications

for Situation V

(Unequal Covariances, Minimal Separation, Three Groups)

Internal Classification

Linear

		Classified Group 1	2	3	Total
Actual	1	53	4	8	65
Group	2	28	11	8	47
	3	23	4	13	40

$P_o = .507$
$P_e = .382$

Quadratic

		Classified Group 1	2	3	Total
Actual	1	51	6	8	65
Group	2	16	25	6	47
	3	13	6	21	40

$P_o = .638$
$P_e = .361$

External Classification

Linear

		Classified Group 1	2	3	Total
Actual	1	44	13	8	65
Group	2	36	1	10	47
	3	23	11	6	40

$P_o = .336$
$P_e = .382$

Quadratic

		Classified Group 1	2	3	Total
Actual	1	33	15	17	65
Group	2	28	8	11	47
	3	20	14	6	40

$P_o = .309$
$P_e = .362$

Table 7

Frequencies of Classifications

for Situation VI

(Unequal Covariances, Nonminimal Separation, Two Groups)

Internal Classification

Linear

	Classified Group 1	2	Total
Actual 1	11	15	26
Group 2	8	192	200

$P_o = .898$

$P_e = .820$

Quadratic

	Classified Group 1	2	Total
Actual 1	26	0	26
Group 2	2	198	200

$P_o = .991$

$P_e = .790$

External Classification

Linear

	Classified Group 1	2	Total
Actual 1	5	21	26
Group 2	9	191	200

$P_o = .867$

$P_e = .838$

Quadratic

	Classified Group 1	2	Total
Actual 1	26	0	26
Group 2	2	198	200

$P_o = .991$

$P_e = .790$

Table 8

Frequencies of Classifications

for Situation VII

(Unequal Covariances, Nonminimal Separation, Three Groups)

Internal Classification

Linear

		Classified Group 1	2	3	Total
Actual	1	137	10	30	177
Group	2	40	13	22	75
	3	37	2	161	200

$P_o = .688$
$P_e = .403$

Quadratic

		Classified Group 1	2	3	Total
Actual	1	137	10	30	177
Group	2	14	48	13	75
	3	28	13	159	200

$P_o = .761$
$P_e = .379$

External Classification

Linear

		Classified Group 1	2	3	Total
Actual	1	129	12	36	177
Group	2	44	9	22	75
	3	41	4	155	200

$P_o = .648$
$P_e = .403$

Quadratic

		Classified Group 1	2	3	Total
Actual	1	115	24	38	177
Group	2	35	16	24	75
	3	37	21	142	200

$P_o = .604$
$P_e = .384$

Table 9

Frequencies of Classifications

Using Nine Measures of Situation VII

(Unequal Covariances, Nonminimal Separation, Three Groups)

Internal Classification

Linear

		Classified Group 1	2	3	Total
Actual	1	135	4	38	177
Group	2	39	11	25	75
	3	38	3	159	200

$P_o = .675$
$P_e = .408$

Quadratic

		Classified Group 1	2	3	Total
Actual	1	134	9	34	177
Group	2	31	24	20	75
	3	35	7	158	200

$P_o = .699$
$P_e = .395$

External Classification

Linear

		Classified Group 1	2	3	Total
Actual	1	132	7	38	177
Group	2	42	7	26	75
	3	41	4	155	200

$P_o = .650$
$P_e = .407$

Quadratic

		Classified Group 1	2	3	Total
Actual	1	121	15	41	177
Group	2	40	12	23	75
	3	39	10	151	200

$P_o = .628$
$P_e = .397$